Ejercicios de Bootstrap

Asignatura	Métodos Avanzados en Estadística
Correo	gloria.valle@estudiante.uam.es
Día	@October 25, 2021
Estudiante	Gloria del Valle Cano
Tema	Tema 1

Ejercicio 2

Sea $X_1 ... X_n \overset{iid}{\sim} F$ con esperanza μ y varianza σ^2, y sea $X_1^* ... X_n^* \overset{iid}{\sim} F_n$. Calcular las siguientes cantidades:

a) $E_{Fn}(\bar{X}^*) \coloneqq E(\bar{X}^*|X_1 ... X_n)$

La distribución bootstrap se puede representar como:

$\bar{x}^*$	x_1	$\cdots$	x_n
$P(X^* = x^*)$	1/n	$\cdots$	1/n

Teniendo en cuenta que la esperanza es un operador lineal, podemos intuir que $\bar{X}^*$ es independiente solo si conoce $X_1 ... X_n$.

$$E(\bar{X}^*|X_1 ... X_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_1^*) = E(X_1^*) = \bar{x}$$

b) $E_F(\bar{X}_n^*)$

En este caso se pregunta por la esperanza sin estar condicionada a la muestra original, lo que lo podemos interpretar como el comportamiento de la media (o cómo varía) en función de la muestra original, lo que es interpretable como la esperanza en sí misma (Ley de las Esperanzas Iteradas), $E[Y] = E[E[Y|X]]$, donde se toma la primera esperanza (interna) con respecto a la variable aleatoria Y y la segunda esperanza (externa) con respecto a X.

$$E(\bar{X}_n^*) = E(E(\bar{X}^*|X_1 ... X_n)) = E(\bar{x}) = \mu$$

c) $Var_{Fn}(\bar{X}^*) \coloneqq Var(\bar{X}|X_1 ... X_n)$

Como la varianza no es un operador lineal:

$$\begin{aligned} Var(\bar{X}^*|X_1 ... X_n) &= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i^*|X_1 ... X_n) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_1^*|X_1 ... X_n) \\ &= \frac{1}{n} Var(X_1^*|X_1 ... X_n) \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n} \\ \left(\text{multiplicando y dividiendo por } \frac{n-1}{n-1}\right) &= \frac{n-1}{n^2} s^2 \end{aligned}$$

Tomando como varianza muestral:

$$s_{n-1}^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$$

d) $Var_F(\bar{X}^*)$

Según la Ley de la Varianza Total, dadas dos variables aleatorias X e Y en el mismo espacio de probabilidad y siendo la varianza de Y finita, entonces $Var[Y] = E[Var[Y|X]] + Var[E[Y|X]]$, donde se toma la primera esperanza/varianza (interna) con respecto a la variable aleatoria Y y la segunda esperanza/varianza (externa) con respecto a X. Por lo tanto, la varianza no condicional es igual a la esperanza de la varianza condicional más la varianza de la esperanza condicional. De esta manera, se toma el ejercicio partiendo de los resultados previamente obtenidos.

$$
\begin{aligned}
Var_F(\bar{X}_n^*) &= E[Var(\bar{X}^*|X_1...X_n)] + Var[E(\bar{X}^*|X_1...X_n)] \\
&= E[\frac{n-1}{n^2}s^2] + Var[\bar{x}] \\
&= \frac{n-1}{n^2}E[s^2] + \frac{1}{n}Var(X_1^*) \\
&= \frac{n-1}{n^2}E[s^2] + \frac{\sigma^2}{n} \\
&= \frac{n-1}{n^2}\sigma^2 + \frac{\sigma^2}{n} \\
&= \frac{\sigma^2}{n}\Big(\frac{n-1}{n} + 1\Big) \\
&= \frac{\sigma^2}{n}\Big(\frac{n-1}{n} + \frac{n}{n}\Big) \\
&= \underbrace{\frac{\sigma^2}{n}}_{Var(\bar{x})}\Big(2 - \frac{1}{n}\Big) \\
&= Var(\bar{x})\Big(2 - \frac{1}{n}\Big)
\end{aligned}
$$

En términos prácticos si $n \to \infty$ entonces $Var(\bar{X}^*) \approx 2\ Var(\bar{x})$.

Ejercicio 7

Sea F una distribución con media μ, varianza σ^2 y coeficiente de asimetría $\gamma = E_F[(X-\mu)^3]/\sigma^3$. Genera $R = 1000$ muestras de observaciones iid $X_1...X_n$ con $X_i \equiv N(0,1)$ para $n = 100$. Para cada una de ellas, calcula tres intervalos de confianza bootstrap de nivel 95% para γ usando el método híbrido, el método normal y el método percentil. Determina el porcentaje de intervalos que contienen al parámetro en cada caso. Repite el ejercicio con muestras procedentes de una distribución exponencial de parámetro $\lambda = 1$.

Se facilita a la derecha el código completo en R de los ejercicios realizados para cada método (ver HTML). A continuación se exponen el razonamiento y los resultados obtenidos.

Para cada método se han definido los siguientes intervalos de confianza:

1️⃣ Método híbrido

$$IC_{1-\alpha}(\theta) = \big[\hat{\theta} - n^{-1/2}H_n^{-1}(1-\alpha/2),\ \hat{\theta} - n^{-1/2}H_n^{-1}(\alpha/2)\big].$$

2️⃣ Método normal

$$IC_{1-\alpha}(\theta) = [\hat{\theta} \mp z_{\alpha/2}\ et_{boot}(\hat{\theta})],$$

donde $et_{boot}(\hat{\theta})$ es un estimador de la desviación típica (error típico) de $\hat{\theta}$.

3️⃣ Método percentil

$$IC_{1-\alpha}(\theta) = [\hat{H}_n^{-1}(\alpha/2),\ \hat{H}_n^{-1}(1-\alpha/2)]$$

Código (R)

💻 Método híbrido

💻 Método normal

💻 Método percentil

Definimos los parámetros en común y una función para calcular el coeficiente de asimetría dada una muestra.

```
set.seed(100)
# Parámetros
R <- 1000
n <- 100
m <- 100
alfa <- 0.05
```

```
coef_asimetria <- function(muestra)
{
  media <- mean(muestra)
  dt <- sd(muestra)
  mean((muestra - media)^3)/dt^3
}
```

Distribución normal $N(\mu, \sigma)$

Se obtiene mediante la función `rnorm` con media 0 y varianza 1.

Método híbrido

```
theta <- 0

# Cálculo de los intervalos
acierto <- NULL
intervalo <- NULL

for (i in 1:m){

  muestra_original <- rnorm(n, 0, 1) + theta
  gamma_original <- coef_asimetria(muestra_original)

  muestras_bootstrap <- sample(muestra_original, n*R, rep = TRUE)
  muestras_bootstrap <- matrix(muestras_bootstrap, nrow = n)

  gamma_bootstrap <- apply(muestras_bootstrap, 2, coef_asimetria)
  T_bootstrap <- sqrt(n) * (gamma_bootstrap - gamma_original)

  ic_min <- gamma_original -  quantile(T_bootstrap, 1-alfa/2)/sqrt(n)
  ic_max  <- gamma_original -  quantile(T_bootstrap, alfa/2)/sqrt(n)
  intervalo <- rbind(intervalo, c(ic_min, ic_max))
  acierto <- c(acierto, ic_min < theta & ic_max > theta)
}
```

Mediante este método se obtiene un 93% de acierto bajo la distribución normal.

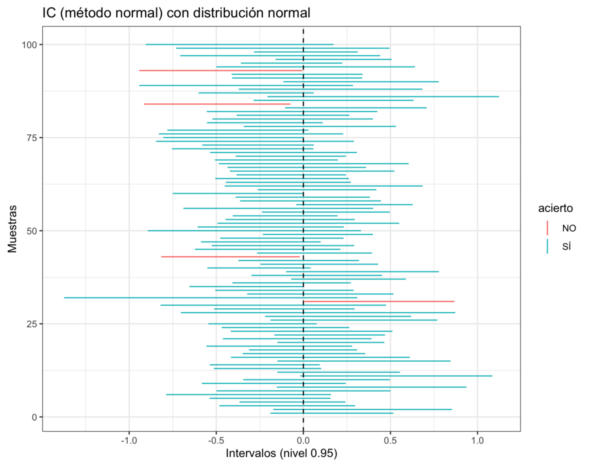
Distribución exponencial $Exp(\lambda)$

Se obtiene mediante la función `rexp` con $\lambda = 1$.

Método híbrido

```
theta <- 2

# Cálculo de los intervalos
acierto <- NULL
intervalo <- NULL

for (i in 1:m){

  muestra_original <- rexp(n, 1) + theta
  gamma_original <- coef_asimetria(muestra_original)

  muestras_bootstrap <- sample(muestra_original, n*R, rep = TRUE)
  muestras_bootstrap <- matrix(muestras_bootstrap, nrow = n)

  gamma_bootstrap <- apply(muestras_bootstrap, 2, coef_asimetria)
  T_bootstrap <- sqrt(n) * (gamma_bootstrap - gamma_original)

  ic_min <- gamma_original -  quantile(T_bootstrap, 1-alfa/2)/sqrt(n)
  ic_max  <- gamma_original -  quantile(T_bootstrap, alfa/2)/sqrt(n)
  intervalo <- rbind(intervalo, c(ic_min, ic_max))
  acierto <- c(acierto, ic_min < theta & ic_max > theta)
}
```

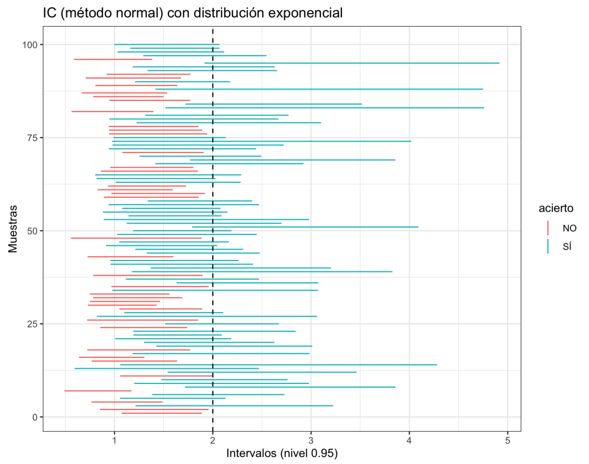
Mediante este método se obtiene un 60% de acierto bajo esta distribución.



Método normal

```
theta <- 0

# Cálculo de los intervalos
acierto <- NULL
intervalo <- NULL

for (i in 1:m){

  muestra_original <- rnorm(n, 0, 1) + theta
  gamma_original <- coef_asimetria(muestra_original)

  muestras_bootstrap <- sample(muestra_original, n*R, rep = TRUE)
  muestras_bootstrap <- matrix(muestras_bootstrap, nrow = n)

  gamma_bootstrap <- apply(muestras_bootstrap, 2, coef_asimetria)
  T_bootstrap <- sqrt(n) * (gamma_bootstrap - gamma_original)

  ic_min <- gamma_original + qnorm(alfa/2, 0, 1)*sd(gamma_bootstrap)
  ic_max  <- gamma_original + qnorm(1-alfa/2, 0, 1)*sd(gamma_bootstrap)

  intervalo <- rbind(intervalo, c(ic_min, ic_max))
  acierto <- c(acierto, ic_min < theta & ic_max > theta)
}
```

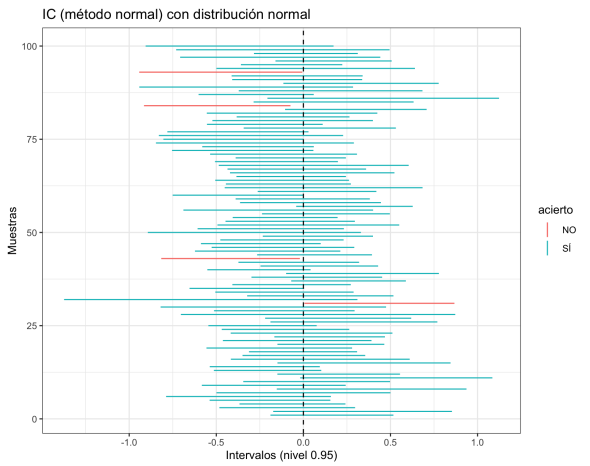
Mediante este método se obtiene un 96% de acierto bajo la distribución normal.



Método normal

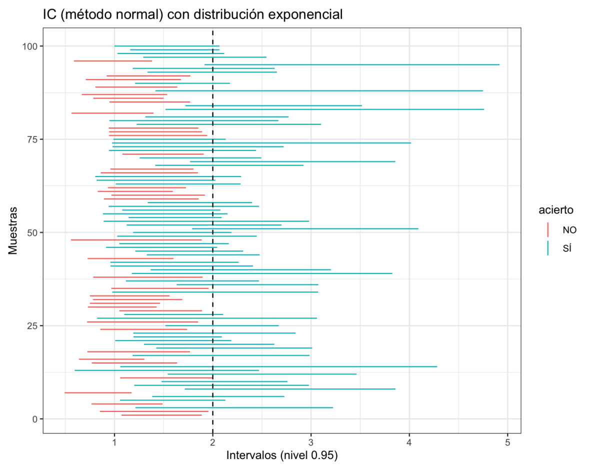
```
theta <- 2

# Cálculo de los intervalos
acierto <- NULL
intervalo <- NULL

for (i in 1:m){

  muestra_original <- rexp(n, 1) + theta
  gamma_original <- coef_asimetria(muestra_original)

  muestras_bootstrap <- sample(muestra_original, n*R, rep = TRUE)
  muestras_bootstrap <- matrix(muestras_bootstrap, nrow = n)

  gamma_bootstrap <- apply(muestras_bootstrap, 2, coef_asimetria)
  T_bootstrap <- sqrt(n) * (gamma_bootstrap - gamma_original)

  ic_min <- gamma_original + qnorm(alfa/2, 0, 1)*sd(gamma_bootstrap)
  ic_max  <- gamma_original + qnorm(1-alfa/2, 0, 1)*sd(gamma_bootstrap)

  intervalo <- rbind(intervalo, c(ic_min, ic_max))
  acierto <- c(acierto, ic_min < theta & ic_max > theta)
}
```

Mediante este método se obtiene un 63% de acierto bajo esta distribución.



Método percentil

```
theta <- 0

# Cálculo de los intervalos
acierto <- NULL
intervalo <- NULL

for (i in 1:m){

  muestra_original <- rnorm(n, 0, 1)
  gamma_original <- coef_asimetria(muestra_original)

  muestras_bootstrap <- sample(muestra_original, n*R, rep = TRUE)
```

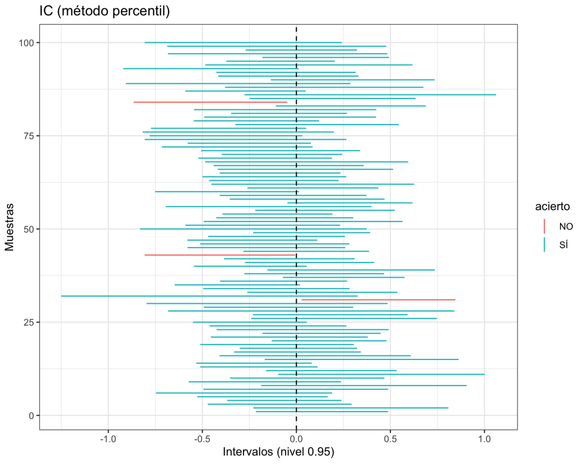


Método percentil

```
theta <- 2

# Cálculo de los intervalos
acierto <- NULL
intervalo <- NULL

for (i in 1:m){

  muestra_original <- rexp(n, 1)
  gamma_original <- coef_asimetria(muestra_original)

  muestras_bootstrap <- sample(muestra_original, n*R, rep = TRUE)
```

```
  muestras_bootstrap <- matrix(muestras_bootstrap, nrow = n)

  gamma_bootstrap <- apply(muestras_bootstrap, 2, coef_asimetria)

  ic_min <- quantile(gamma_bootstrap, alfa/2)
  ic_max  <- quantile(gamma_bootstrap, 1-alfa/2)
  intervalo <- rbind(intervalo, c(ic_min, ic_max))
  acierto <- c(acierto, ic_min < theta & ic_max > theta)
}
```

Mediante este método se obtiene un 97% de acierto bajo la distribución normal.




```
  muestras_bootstrap <- matrix(muestras_bootstrap, nrow = n)

  gamma_bootstrap <- apply(muestras_bootstrap, 2, coef_asimetria)

  ic_min <- quantile(gamma_bootstrap, alfa/2)
  ic_max  <- quantile(gamma_bootstrap, 1-alfa/2)
  intervalo <- rbind(intervalo, c(ic_min, ic_max))
  acierto <- c(acierto, ic_min < theta & ic_max > theta)
}
```

Mediante este método se obtiene un 59% de acierto bajo esta distribución.